TEXTURE SIMILARITY METRICS APPLIED TO HEVC INTRA PREDICTION

Karam Naser, Vincent Ricordel, Patrick Le Callet

TEXTURE SIMILARITY METRICS APPLIED TO HEVC INTRA PREDICTION

Karam Naser, Vincent Ricordel and Patrick Le Callet

LUNAM University, University of Nantes, IRCCyN UMR CNRS 6597
Polytech Nantes, Rue Christian Pauc BP 50609 44306 Nantes Cedex 3, France
karam.naser; vincent.ricordel; patrick.le-callet
@univ-nantes.fr

ABSTRACT

Textures represent regions of homogeneous properties (e.g. color, directionality, edge distribution etc.). They cover wide areas of the visual scene with less importance as compared to structures.

The typical image/video encoders aim at optimizing the bitrate within a certain distortion level. The distortion is usually measured via comparing pixel values, which could highly deviates from the perceived distortions especially for texture components. In this paper, we review and verify our recent work on using texture similarity metrics as a measure of perceived distortion in HEVC, and optimize encoder accordingly. Experimental results reveal the same findings in [1] [2] that when a texture similarity metric is used, the visual quality of the decoded textures is improved as well as the rate-similarity performance.

1. INTRODUCTION

Visual signal is a rich source of information. It consists of different components and segments. Among them, textures are special part that are well understood in terms of perception, representation and characterization [3].

Textures are visually less important as compared to structures. The structures convey the semantic of the scene while textures can be considered as a visual enrichment component.

Texture similarity has been studied quite well in the problem of similar texture retrieval. Different studies showed that the metrics based on comparing pixel values (such as PSNR) perform very poorly in retrieving similar textures. Other metrics, which are either based on comparing pixel distributions [4], subband statistics [5], or combination of both [6] can perform much more better (details in [7]).

Video coding standards, such as HEVC, aim at minimizing the bitrate within a certain distortion level. The distortion is typically expressed in terms of pixel differences. As mentioned before, the pixel comparison of texture images does not proportionally reflect the amount of perceived distortion. This reason has motivated us to investigate about the practicality of using different types of texture similarity metrics in [1] and [2]. It was shown that those metrics generally improve the visual quality of the decoded textures, as well as the rate-similarity performance.

In this paper, we review our recent work on applying the texture similarity metrics in HEVC, where texture similarity metrics were used as a distortion measure inside the encoder cost function for selecting the best intra prediction mode and block partitioning. Besides, it verifies the results using different dataset.

The rest of the paper is organized as follows: Section 2 gives an overview of the texture similarity metrics used in this work. Section 3 presents the procedure used to evaluate each metric. In Section 4, the experimental results are provided and discussed. In Section 5, the verification procedure for the result is described. The discussion and conclusion are given in Section 6.

2. OVERVIEW OF TEXTURE SIMILARITY METRICS

Texture similarity metrics exist in various forms. Some of them compare the statistics of textures in the spatial domain and others in the subband frequency domain (details can be found in [7]. In our work, we considered two metrics (STSIM and LRI) as being new and successful texture similarity metrics.

2.1. STSIM

The Structural Texture Similarity metric (STSIM) is highly inspired by SSIM and its frequency domain implementation (CW-SSIM). However, this metric is meant for texture similarity rather than image quality assessment. It was presented in [8] and further improved in [9]. STSIM is based on comparing a set of statistics in the subband decomposition. These statistics consist of mean, standard deviation, and horizontal and vertical auto-correlation of each subband. Beside that, it computes also the cross correlations between subbands with the same scale or with the same orientation. This set of statistics provides a solid description of a given texture and thus can well characterize the similarity between two textures.

2.2. LRI

The Local Radius index LRI is a successive to STSIM. LRI is much less computationally expensive as compared to STSIM and performs better in the context of similar texture retrieval [6]. It computes a local index for each pixel in the spatial domain, beside that, it computes the local binary pattern, standard deviation of each subband in the subband frequency domain and an intensity penalization term. Thus it is a combination between the analysis in frequency decomposition and spatial domain.

3. PERFORMANCE EVALUATION

We implemented both STSIM and LRI, and integrated them in the HEVC reference encoder. These metrics were used inside the cost function for selecting the best intra prediction mode and block partitioning. In the following subsections, the details are given:

3.1. Intra Mode Selection in HEVC reference Encoder

HEVC defines 35 possible prediction modes (33 directional modes, DC and Planar mode)[10]. The prediction block starts from the size of 64x64 and can be split in a quadtree manner up to 4x4. The best prediction mode and split are determined by a cost function which combines both the rate and distortion.

In the reference implementation, this full optimization process is avoided as it is time consuming. Instead HEVC pre-selects 3 most probable prediction modes. These modes are the ones among the 35 that minimize the Sum of Absolute Transformed Difference (SATD) and the rate of coding the prediction information. Using these 3 modes, it tries then encoding and splitting until it finds the best combination. The used distortion in this case is the Sum of Squared Difference (SSD).

3.2. Replacing HEVC Distortion Metric

In our work, we replaced the SATD and SSD of HEVC reference encoder by texture similarity metrics (STSIM and LRI). These metrics were adapted and scaled (details in [1][2]) to match the range of SATD and SSD.

4. EXPERIMENTS AND RESULTS

We have experimented the use of STSIM and LRI in HEVC for coding static textures. We used Brodazt textures downloaded from USC-SIPI dataset [11]. This contains 13 different gray scale textures which are extensively used in textures analysis for engineering and psychophysical experiments. We used HM 9.0 [12] as a host encoder. In the following subsections, we provide the details of each experiment.

4.1. Quality of the Decoded Textures

To understand the impact of using different metrics in HEVC, Fig. 1 gives examples of the decoded textures when using the same quantization parameter (QP).

In the first row of Fig. 1, we see an example of encoding a highly structured texture. In high compression scenario, the texture looses most of its details when the default HEVC metrics are used. This is because many blocks are replaced by DC values. Using any of the similarity metrics can retain the overall structure of the texture. One can also notice that there exists many wrong directions, but the overall quality is much more better.

Another example is shown in the second row. In this example, the effect of wrong prediction direction is more clear when LRI is used. On the other hand, the right part of the texture is completely eliminated when the default metrics are used.

The third example (last row of Fig. 1) of bubbles is a good example of high deviation from pixel fidelity when LRI or STSIM are used. We can see that the bubbles don't appear closed anymore and many direction appear which where not available in the original image. But overall, the decoded textures appear more pleasant when LRI or STSIM is used.

4.2. Rate Distortion Analysis

The rate distortion analysis is usually carried out using PSNR as a distortion measure. In our approach, PSNR is avoided as it is based on pixel difference, which is far away from the goal of this work. For this purpose, we sought another metric that is specifically designed for textures.

We used a texture similarity metric [13] which is based on comparing features of textures in the wavelet domain. These features correspond to the mean and standard of the subband images obtained using Gabor filters. This metric often provides close by performance as compared to LRI and STSIM (cf. [9]) in terms of retrieval rate. The metric was downloaded from the authors website and used as a distortion metrics in our work.

By calculating the distance measured by this metric to the original texture for different compression levels, we obtained the curves shown in Fig 2. We observe that in most cases, LRI and STSIM provides better score than the default metrics in the low rate region. For high rate region, no gain is achieved.

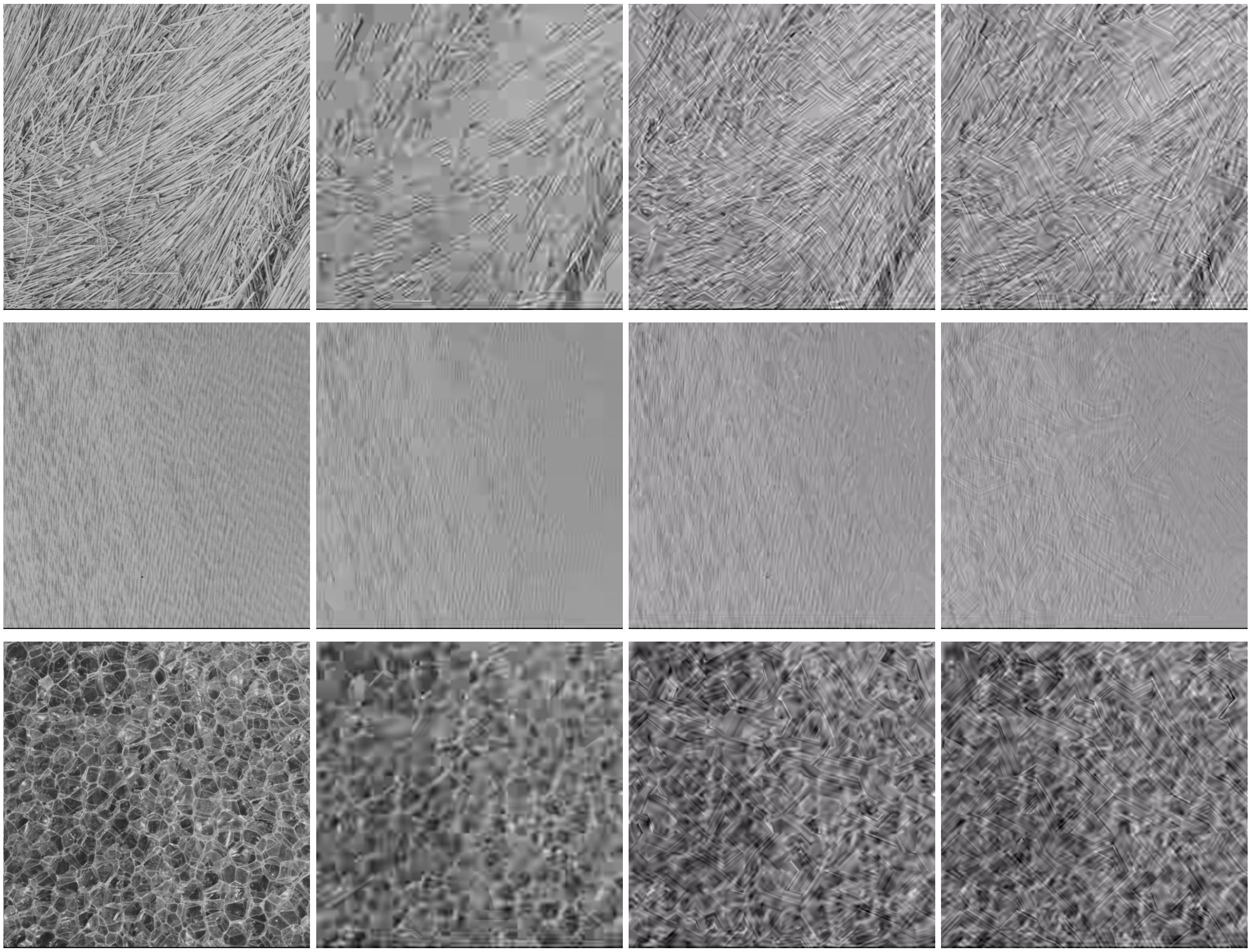

Fig. 1. Examples of decoded textures using the same QP. From left to right: Original texture, compressed using HEVC default metrics, using STSIM and using LRI.

4.3. Encoder Behavior Analysis

More analysis was done to understand the effects of using the perceptual metrics on the prediction mechanism. For this, we measured the frequency of splitting depths as a function of the quantization parameter. The corresponding histograms are shown in Fig. 3. The splitting depth of zero means that the prediction block has its maximal size (64x64). Increasing the splitting depth by one corresponds to partition the block into four sub-blocks. The histograms in Fig. 3 were scaled by the number of (4x4) blocks that each splitting has. This was done to have a fair comparison between splitting depths as each splitting occupies different areas of the frame. One can observe from these histograms that when the default metrics are used, the encoder uses small prediction blocks for low compression (low QP) to better approximate the input signal. For high compression, it tries to approximate large prediction blocks (mostly with DC values) to have better compression. The behavior is totally changes when LRI or STSIM is used. The encoder behavior does not change much as the compression changes. It selects always large block sizes to approximate the input signal and small block sizes (less than 16x16) are rarely chosen. The is because these metrics compare statistics of different distributions. For small block sizes, there is always a lack of enough statistics and usually these metrics return a high value of distortion in such a condition.

5. VERIFICATION OF THE RESULTS

To verify the results, we repeated the same experiments using different dataset of textures. This time, we used some textures from QualTex texture dataset [14]. Examples of the decoded textures are shown in Fig. 4. We can clearly see that the fine details of the texture are better preserved when

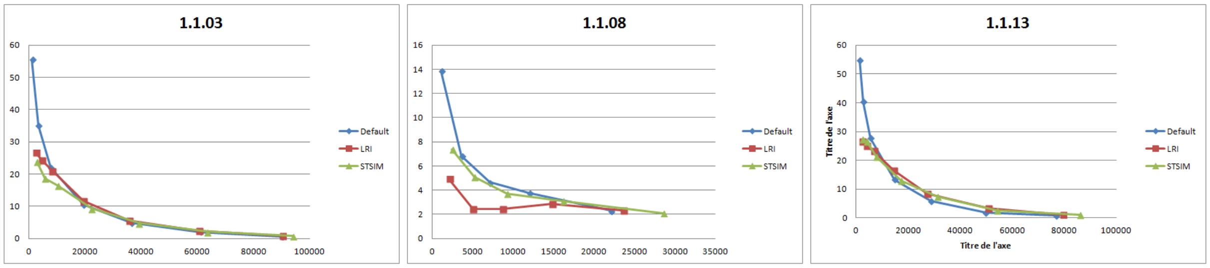

Fig. 2. Rate Distortion (using Gabor distance metric [13]) of the textures shown in Fig. 1. x-axes: Bytes used to encode the texture, y-axes: distance to the original texture. Indexes above each curve correspond to the same naming terminology in the dataset.

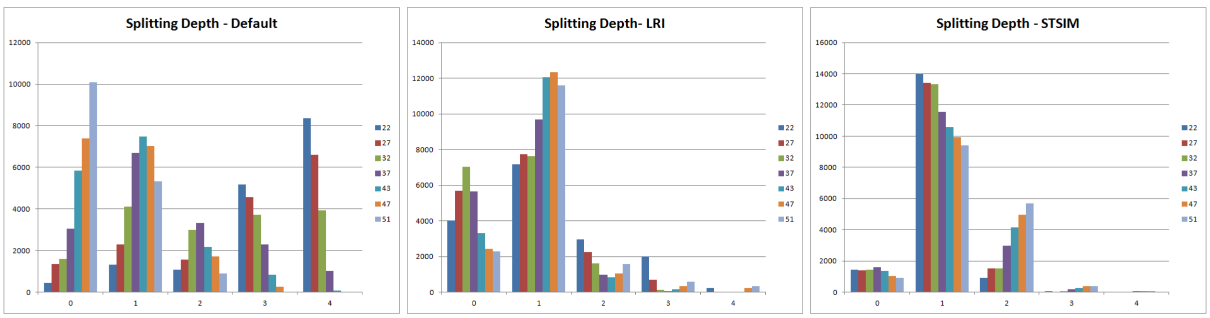

Fig. 3. Histograms of splitting depths vs QP. Each depth is scaled by the number of 4x4 block that it has.

the texture similarity metrics are used, but when using the default metrics, all images look more blurry compared to the original.

The rate similarity curves are shown in Fig. 5. These curve are very much consistent with curves obtained using Brodatz textures (Fig. 2). This indicates clearly that these metrics perform better in low rate scenario.

The encoder behaves similarly in both datasets. As we see in Fig. 6, when STSIM or LRI are used, the encoder uses larger blocks independently from QP, this is for the same reason mentioned previously (c.f. 4.3).

6. DISCUSSION AND CONCLUSION

In this paper, we reviewed our recent work on embedding texture similarity metrics in HEVC intra prediction. Besides, we verified the results using another dataset. The results (using different datasets) are consistent in terms of visual quality, rate distortion curves and encoder behavior.

The texture similarity metrics retain the structural information of the textures, in contrast to the default metrics which try to smooth the contents and replace them by DC values. For severe compression, the decoded textures can have a noisy structure as HEVC intra prediction cannot provide anything better than parallel lines of the directional prediction. Using both metrics, wrong prediction directions might be selected. This is because these metrics are less sensitive to pixel by pixel comparison. LRI, as compared to STSIM, is much less computationally expensive. But it results in more wrong prediction directions than STSIM.

The encoder behaves differently when the similarity metrics are used. It uses mostly large prediction block sizes for all range of compression. This is mainly because in small blocks, there is a lack of enough statistics to compare and the metrics will return high dis-similarity values.

The rate-distortion curves, which were obtained using the texture similarity metrics [13], show that both metrics perform better than the default metric (in low bitrate scenario).

The direct benefit of this approach its compatibility with HEVC standard, which means no modification to the decoder is needed. For the encoder, there is still a room for improvement using rate-distortion optimization techniques, which is left for future work.

As a conclusion, the use of texture similarity metrics can generally give a better visual quality than the default ones. The approach presented in this paper is a kind of features-aware coding, in which the encoder takes into consideration of the signal features during the compression process. However, finding the correct features which match the human perception would be the optimal goal.

Fig. 4. Examples of decoded textures using the same QP. From left to right: Original texture, compressed using HEVC default metrics, using STSIM and using LRI.

7. ACKNOWLEDGMENT

This work was supported by Marie Sklodowska-Curie Initial Training Network under the PROVISION (PeRceptually Optimised VIdeo CompresSION) project bearing Grant Number 608231 and Call Identifier: FP7-PEOPLE-2013-ITN.

8. REFERENCES

[1] K. Naser, V. Ricordel, and P. Le Callet, "Experimenting texture similarity metric STSIM for intra prediction mode selection and block partitioning in HEVC," in *Digital Signal Processing (DSP), 2014 19th International Conference on*. IEEE, 2014, pp. 882–887.

[2] ——, "PERFORMANCE ANALYSIS OF TEXTURE SIMILARITY METRICS IN HEVC INTRA PREDICTION," in *Int. Workshop Video Processing and Quality Metrics for Consumer Electronics*, 2015.

[3] M. Tuceryan and A. K. Jain, "Texture analysis," *The handbook of pattern recognition and computer vision*, vol. 2, pp. 207–248, 1998.

[4] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[5] J. Ehmann, T. Pappas, and D. Neuhoff, "Structural texture similarity metrics for image analysis and retrieval," 2013.

[6] Y. Zhai, D. Neuhoff, and T. Pappas, "Local radius index - a new texture similarity feature," in *Acoustics,*

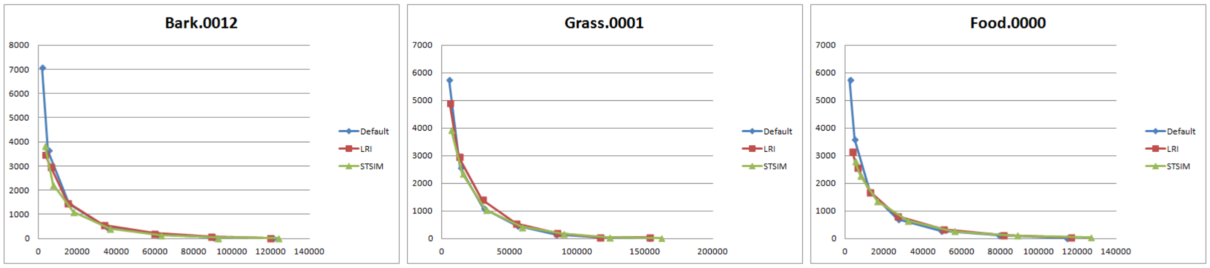

Fig. 5. Rate Distortion (using Gabor distance metric [13]) of the textures shown in Fig. 1. x-axes: Bytes used to encode the texture, y-axes: distance to the original texture. Indexes above each curve correspond to the same naming terminology in the dataset.

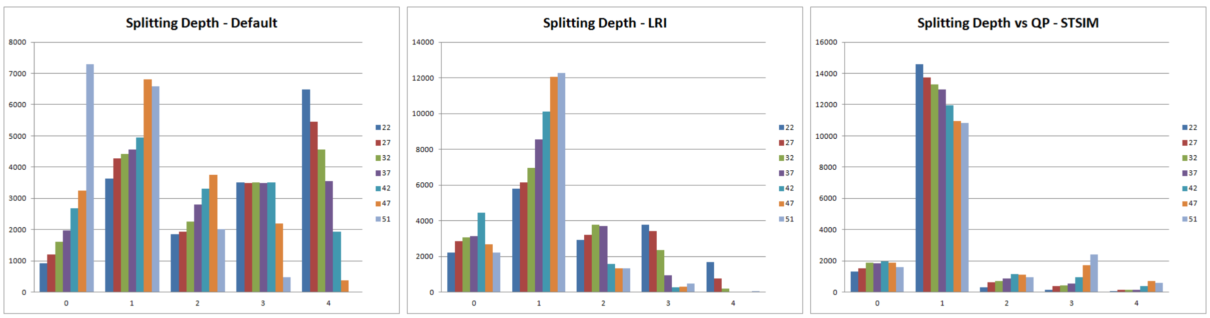

Fig. 6. Histograms of splitting depths vs QP. Each depth is scaled by the number of 4x4 block that it has.

Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, May 2013, pp. 1434–1438.

[7] T. Pappas, D. Neuhoff, H. de Ridder, and J. Zujovic, "Image Analysis: Focus on Texture Similarity," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2044–2057, Sept 2013.

[8] X. Zhao, M. G. Reyes, T. N. Pappas, and D. L. Neuhoff, "Structural texture similarity metrics for retrieval applications," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 1196–1199.

[9] J. Zujovic, T. Pappas, and D. Neuhoff, "Structural Texture Similarity Metrics for Image Analysis and Retrieval," *Image Processing, IEEE Transactions on*, vol. 22, no. 7, pp. 2545–2558, July 2013.

[10] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.

[11] "USC-SIPI Dataset." [Online]. Available: http://sipi.usc.edu/database/

[12] Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG, "HEVC test model 9.0," Tech. Rep.

[13] B. S. Manjunath and W.-Y. Ma, "Texture features for browsing and retrieval of image data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 8, pp. 837–842, 1996.

[14] M. S. Gide and L. J. Karam, "On the assessment of the quality of textures in visual media," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*. IEEE, 2010, pp. 1–5.